

# Supporting Information for Pepsi-SAXS : an adaptive method for rapid and accurate computation of small angle X-ray scattering profiles.

Sergei Grudin <sup>\*a,b,c</sup>, Maria Garkavenko<sup>d</sup> and Andrei Kazennov<sup>d</sup>

<sup>a</sup>Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France

<sup>b</sup>CNRS, LJK, F-38000 Grenoble, France

<sup>c</sup>Inria

<sup>d</sup>Moscow Institute of Physics and Technology, Dolgoprudniy, Russia

## 1 Fitting with partial intensities

To speed-up the calculations of the theoretical scattering intensity curve at different values of  $\delta r$  and  $\delta \rho$ , we re-write it as a sum of partial intensities,

$$I_{theor}(q) = I_{aa}(q) + \delta \rho^2 I_{bb}(q) + G(q, \delta r) I_{cc}(q) + \delta \rho I_{ab}(q) + G(q, \delta r) I_{ac}(q) + G(q, \delta r) \delta \rho I_{bc}(q), \quad (1)$$

where the partial intensities are computed only once as follows,

$$\begin{aligned} I_{aa}(q) &= \sum_{l=0}^L \sum_{m=-l}^l A_{lm}(q)^2 \\ I_{bb}(q) &= \sum_{l=0}^L \sum_{m=-l}^l B_{lm}(q)^2 \\ I_{cc}(q) &= \rho^2 \sum_{l=0}^L \sum_{m=-l}^l C_{lm}(q)^2 \\ I_{ab}(q) &= 2 \sum_{l=0}^L \sum_{m=-l}^l A_{lm}(q) B_{lm}(q) \\ I_{bc}(q) &= -2\rho \sum_{l=0}^L \sum_{m=-l}^l B_{lm}(q) C_{lm}(q) \\ I_{ac}(q) &= -2\rho \sum_{l=0}^L \sum_{m=-l}^l A_{lm}(q) C_{lm}(q). \end{aligned} \quad (2)$$

This allows us to reduce the computational cost of the theoretical scattering intensity curve by a factor of  $O(L^2)$ .

## 2 Fitting with a constant

Some experimental measurements have a systematic error in the determination of the intensity values. To account for this error, we introduce the offset constant  $\kappa$  and re-write the goodness of fit  $\chi^2$  as

$$\chi^2 = \frac{1}{N-1} \sum_j \left( \frac{I_{exp}(q_j) + \kappa - c(\kappa) I_{theor}(q_j)}{\sigma(q_j)} \right)^2, \quad (3)$$

where the scaling factor  $c(\kappa)$  is

$$c(\kappa) = c + \kappa b, \quad (4)$$

---

\*sergei.grudin@inria.fr

with the constant  $b$  given as

$$b = \frac{\sum_j I_{theor}(q_j)/\sigma(q_j)^2}{\sum_j I_{theor}(q_j)^2/\sigma(q_j)^2}. \quad (5)$$

Now we can compute the optimal offset constant  $\kappa$  for each sampled value of  $\delta r$  and  $\delta \rho$  of the theoretical intensity curve  $I_{theor}(q)$  by analytically minimizing the least-square discrepancy  $\chi^2$  as follows,

$$\kappa = \frac{\sum_j (I_{exp}(q_j) - cI_{theor}(q_j))(1 - bI_{theor}(q_j))/\sigma(q_j)^2}{\sum_j (1 - bI_{theor}(q_j))^2/\sigma(q_j)^2}. \quad (6)$$

### 3 Additional Benchmarks

Please note that in these benchmarks we did not restrict the angular range to the default value of  $0.5 \text{ \AA}^{-1}$ .

Table S1: Benchmarking Pepsi-SAXS without spline interpolation (4th column), without contrast of the hydration shell ( $d\rho$  set to  $0 \text{ e/nm}$ , 5th column), with the constant variation of the hydration shell ( $d\rho$  set to the average over the dataset,  $18 \text{ e/nm}$ , 6th column), with a wider search window for the adjustable  $d\rho$  parameter equivalent to the one used in CRY SOL (7th column), and with a negative contrast of the hydration shell (8th column). Experimental data is collected from the BioIsis and SASBDB databases. For each test, we also provide the reference values of  $\chi$  for each profile (3rd column) along with the average values of  $\chi$  over the dataset.

Structure name	Curve name	$\chi$ values					
		Default	No spline	$d\rho = 0 \text{ e/nm}$	$d\rho = 18 \text{ e/nm}$	$d\rho_{max} = 60 \text{ e/nm}$	$d\rho_{min} = -15 \text{ e/nm}$
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	1.39	1.38	1.93	1.56	1.39	1.39
Abscisic acid receptor PYR1	BID_1PYR1P	2.33	2.34	2.63	3.74	2.33	2.33
Rubredoxin	BID_1RBDGP	7.24	7.37	62.84	17.12	7.24	7.24
Superoxide reductase	BID_1SPXGP	1.67	1.67	12.75	4.37	1.69	1.69
Monomeric PF1674	BID_1TSPHP	6.22	6.16	7.27	17.09	6.22	6.22
Endo-1,4-beta-xylanase II	BID_1XYNTP	0.93	0.93	2.78	1.26	0.92	0.92
28 bp DNA	BID_28BPDD	0.56	0.55	1.88	0.76	0.55	0.55
S-adenosylmethionine riboswitch mRNA	BID_2SAMRR	2.46	2.45	2.82	2.46	2.46	2.46
Superoxide dismutase	BID_APSODP	3.87	3.88	4.98	3.95	3.87	3.87
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	2.19	2.19	6.14	2.83	2.19	2.19
Complement C3b-Efb (from S. aureus)	BID_C3bEfP	1.63	1.62	7.60	2.79	1.50	1.50
complement C3b + Efb (staphylococcal)	BID_C3BSAP	0.12	0.12	0.14	0.12	0.12	0.12
Glucose Isomerase	BID_GIKCIP	7.19	7.27	11.56	7.26	7.19	7.19
Glucose Isomerase	BID_GISRUP	3.31	3.34	9.45	3.31	3.31	3.31
Human Regulator of Chromosome Condensation (RCC1)	BID_HRCC1P	1.59	1.59	7.32	3.72	0.98	0.98
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1.85	1.86	6.14	2.68	1.86	1.86
Lysozyme	BID_LYKCIP	9.20	9.21	9.20	10.13	9.20	9.20
Hen egg-white lysozyme	BID_LYSOZP	2.62	2.60	7.57	4.29	2.25	2.25
MnmE in the nucleotide free state	BID_MNME1P	0.83	0.83	1.16	0.83	0.83	0.83
E.c. MnmE/MnmG complex in the nucleotide free state	BID_MnmEGP	1.84	1.84	1.84	1.99	1.84	1.84
A. aeolicus MnmG	BID_MnmG1P	1.44	1.42	2.91	1.49	1.44	1.44
A. aeolicus MnmG + tRNA	BID_MnmG2X	1.75	1.71	1.83	2.21	1.76	1.76
E.coli MnmG + NbMnmG1	BID_MnmG3P	2.03	1.94	11.53	2.18	2.03	2.03
E.c. MnmE/MnmG complex bound to GDP-AlFx	BID_MnmGEP	2.87	2.89	9.42	3.56	2.82	2.82
DNA double-strand break repair protein MRE11	BID_MRERAP	1.29	1.29	6.23	2.21	1.29	1.29
Cu/Zn Superoxide Dismutase	BID_NMSODP	0.97	0.97	4.54	1.91	0.97	0.97
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	0.11	0.11	0.14	0.11	0.11	0.11

Ketoreductase-enoylreductase main	BID_ZGDWKP	1.61	1.61	3.75	1.91	1.61	1.61
Lysozyme C	SASDAC2	1.19	1.18	1.64	1.20	1.19	1.19
Ubiquitin-60S ribosomal protein L40	SASDAQ2	2.77	2.77	4.15	2.90	2.77	2.77
Myoglobin in PBS	SASDAH2	1.99	1.98	3.18	2.03	1.99	1.99
Catalase in PBS	SASDA92	2.92	2.89	3.08	2.97	2.92	2.92
Methyltransferase WbdD	SASDAJ6	1.35	1.34	2.31	1.36	1.35	1.35
Alcohol dehydrogenase in PBS	SASDA52	2.71	2.71	2.71	5.79	2.71	2.71
Calmodulin:peptide complex	SASDAN4	3.48	3.50	11.69	3.53	3.48	3.48
Psi-producing oxygenase A	SASDA45	9.44	9.45	13.06	9.97	9.37	9.37
Cysteine desulfurase IscS dimer	SASDAV6	1.24	1.24	2.52	1.57	1.19	1.19
Factor H CCP modules 12 to 13	SASDA25	1.56	1.56	1.83	2.09	1.56	1.56
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	1.05	1.03	2.47	1.13	1.05	1.05
Complex ComE-comcde	SASDAB7	1.21	1.25	2.58	1.23	1.21	1.21
Complex LytTR-comcde	SASDAC7	1.26	1.19	3.58	1.26	1.26	1.26
Myomesin-1	SASDAK5	1.82	1.82	2.75	2.06	1.82	1.82
IcsS, IcsU and CyaY dimeric complex	SASDA27	2.55	2.55	2.83	2.70	2.55	2.55
Iron-sulfur cluster assembly scaffold	SASDAW6	1.21	1.21	1.45	1.32	1.21	1.21
IscU monomer							
Geminin:Cdt1 2:1 heterotrimer	SASDAV3	2.60	2.60	7.11	4.15	1.78	1.78
Geminin:Cdt1 4:2 heterohexamer	SASDAW3	2.94	2.94	2.94	4.57	2.94	2.94
apo XMRV RT	SASDAV5	1.12	1.12	1.12	1.22	1.12	1.12
XMRV RT + DNA/RNA hybrid	SASDAW5	0.91	0.91	0.92	0.96	0.91	0.91
Protein CyaY monomer	SASDAX6	1.18	1.18	2.11	1.46	1.06	1.06
Annexin-A4	SASDAJ5	4.78	4.80	5.85	4.78	4.78	4.78
Pyruvate decarboxylase	SASDAX2	0.83	0.83	0.89	0.91	0.83	0.83
Average		2.42	2.42	5.71	3.31	2.37	2.39

Table S2: Benchmarking Pepsi-SAXS with different search grids for the fitting parameters  $\delta r$  and  $\delta \rho$  (4th and 5th columns), with a twice wider search window for the adjustable  $dr_0$  parameter, without special form factors for charged and resonance groups (7th column), and without accounting for the hydrogen atoms (8th column). Experimental data is collected from the BioIsis and SASBDB databases. For each test, we also provide the reference values of  $\chi$  for each profile (3rd column) along with the average values of  $\chi$  over the dataset.

Structure name	Curve name	Default	$\chi$ values				
			10×10 grid	1,000×1,000 grid	$dr_0 = \pm 10\%$	No resonance groups	No explicit hydrogens
Rab1 adenylation (AMPylation) protein	BID_1DRRAP	1.39	1.39	1.39	1.39	1.38	1.41
Abscisic acid receptor PYR1	BID_1PYR1P	2.33	2.35	2.33	2.33	2.53	4.99
Rubredoxin	BID_1RBDGP	7.24	7.34	7.23	7.24	7.25	8.04
Superoxide reductase	BID_1SPXGP	1.67	4.00	1.67	1.67	3.58	5.85
Monomeric PF1674	BID_1TSPHP	6.22	7.98	6.22	6.22	6.12	7.52
Endo-1,4-beta-xylanase II	BID_1XYNTP	0.93	0.97	0.93	0.93	0.93	0.97
28 bp DNA	BID_28BPDD	0.56	0.56	0.56	0.56	0.56	0.56
S-adenosylmethionine riboswitch mRNA	BID_2SAMRR	2.46	2.46	2.46	2.42	2.46	2.51
Superoxide dismutase	BID_APSODP	3.87	3.95	3.86	3.87	3.95	3.91
Ubiquitin-like modifier-activating enzyme ATG7 C-terminal domain	BID_ATG7CP	2.19	2.19	2.19	2.19	2.24	2.15
Complement C3b-Efb (from <i>S. aureus</i> )	BID_C3bEfP	1.63	1.63	1.63	1.63	1.64	2.09
complement C3b + Efb (staphylococcal)	BID_C3BSAP	0.12	0.12	0.12	0.12	0.12	0.12
Glucose Isomerase	BID_GIKCIP	7.19	7.25	7.19	7.19	7.47	7.16
Glucose Isomerase	BID_GISRUP	3.31	3.42	3.30	3.40	4.15	4.00
Human Regulator of Chromosome Condensation (RCC1)	BID_HRCC1P	1.59	1.76	1.59	1.59	1.54	2.72
Immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR3	BID_LAR12P	1.85	1.94	1.85	1.85	1.85	1.98
Lysozyme	BID_LYKCIP	9.20	9.20	9.20	8.80	9.20	10.41
Hen egg-white lysozyme	BID_LYSOZP	2.62	2.66	2.62	2.62	2.29	2.71
MnmE in the nucleotide free state	BID_MNME1P	0.83	0.83	0.83	0.83	0.82	0.83
E.c. MnmE/MnmG complex in the nucleotide free state	BID_MnmEGP	1.84	1.86	1.84	1.83	1.85	1.85
A. aeolicus MnmG	BID_MnmG1P	1.44	1.44	1.44	1.44	1.44	1.50
A. aeolicus MnmG + tRNA	BID_MnmG2X	1.75	1.78	1.75	1.75	1.78	1.75
E.coli MnmG + NbMnmG1	BID_MnmG3P	2.03	2.06	2.03	2.03	2.05	2.14
E.c. MnmE/MnmG complex bound to GDP-AlFx	BID_MnmGEP	2.87	2.87	2.87	2.87	2.85	2.89
DNA double-strand break repair protein MRE11	BID_MRERAP	1.29	1.33	1.29	1.29	1.46	1.67
Cu/Zn Superoxide Dismutase	BID_NMSODP	0.97	1.01	0.97	0.98	0.98	1.34
Interleukin (IL)-33 with primary receptor ST2	BID_ST2ILP	0.11	0.11	0.11	0.11	0.11	0.11
Ketoreductase-enoylreductase didomain	BID_ZGDWKP	1.61	1.79	1.61	1.61	1.88	1.97
Lysozyme C	SASDAC2	1.19	1.19	1.19	1.19	1.18	1.19
Ubiquitin-60S ribosomal protein L40	SASDAQ2	2.77	2.84	2.77	2.77	2.79	2.86
Myoglobin in PBS	SASDAH2	1.99	2.08	1.99	1.99	1.99	2.02
Catalase in PBS	SASDA92	2.92	2.92	2.92	2.92	2.95	2.90
Methyltransferase WbdD	SASDAJ6	1.35	1.36	1.35	1.35	1.38	1.37
Alcohol dehydrogenase in PBS	SASDA52	2.71	2.71	2.70	2.71	2.94	2.98
Calmodulin:peptide complex	SASDAN4	3.48	3.70	3.48	3.48	3.61	3.64
Psi-producing oxygenase A	SASDA45	9.44	9.48	9.44	9.44	9.38	9.34
Cysteine desulfurase IscS dimer	SASDAV6	1.24	1.24	1.24	1.24	1.25	1.33
Factor H CCP modules 12 to 13	SASDA25	1.56	1.57	1.56	1.56	1.61	1.60
Heterotetramer of histidine protein kinase and response regulator	SASDAA7	1.05	1.05	1.05	1.05	1.05	1.06



Complex ComE-comcde	SASDAB7	1.21	1.22	1.21	1.21	1.20	1.25
Complex LytTR-comcde	SASDAC7	1.26	1.26	1.26	1.26	1.26	1.34
Myomesin-1	SASDAK5	1.82	1.82	1.82	1.82	1.76	1.74
IcsS, IcsU and CyaY dimeric complex	SASDA27	2.55	2.55	2.55	2.55	2.57	2.63
Iron-sulfur cluster assembly scaffold	SASDAW6	1.21	1.21	1.21	1.21	1.21	1.55
IscU monomer							
Geminin:Cdt1 2:1 heterotrimer	SASDAV3	2.60	2.61	2.60	2.61	2.61	3.34
Geminin:Cdt1 4:2 heterohexamer	SASDAW3	2.94	2.94	2.94	2.94	2.96	3.10
apo XMRV RT	SASDAV5	1.12	1.12	1.12	1.12	1.11	1.10
XMRV RT + DNA/RNA hybrid	SASDAW5	0.91	0.91	0.91	0.91	0.91	0.91
Protein CyaY monomer	SASDAX6	1.18	1.19	1.18	1.18	1.17	1.24
Annexin-A4	SASDAJ5	4.78	4.78	4.78	4.78	4.78	4.77
Pyruvate decarboxylase	SASDAX2	0.83	0.83	0.83	0.83	0.83	0.84
Average		2.42	2.53	2.42	2.41	2.49	2.73

Figure S1: Comparison of the experimental scattering profiles with the modelled ones computed by Pepsi-SAXS, CRYSOL and FoXS on a log scale. For each of the profiles residuals on a linear scale are plotted. Experimental data is collected from the BioIsis and SASBDB databases.







































